# Analyzing NBA Player Value Through Network Analysis

Matt Gentile

*Colgate University*

December 17, 2019

MATH 482 - Seminar: Applied Mathematics

# INTRODUCTION

"The strength of the team is each individual member.
The strength of each member is the team."

*-Phil Jackson*

Phil Jackson won 11 National Basketball Association (NBA) championships as a head coach, more than anyone else in the history of the league. His intricate knowledge of how to maximize his roster's talent, a key to building two of the most dominant dynasties ever (Chicago Bulls in the 1990s and Los Angeles Lakers in early 2000s), earned him the nickname the "Zen Master." Jackson's quote pulls at one of the most fundamental and cliché attributes of basketball: it's a team sport.

Anyone who understands the game of basketball would agree that the game is much more intricate than simply throwing the ball at the basket in hopes of scoring. Players can affect the ebbs and flows of the game in so many various ways that understanding how a player singularly impacts his team is very difficult. Putting this understanding into a number that can be compared against other players is an even bigger challenge.

The goal of this study is to do exactly that. The goal of every NBA team is to win more games and, ultimately, win more championships. Teams should not care about how many players on their team average 20+ points per game or how many players shoot over 40% on three-point field goals. Teams should be concerned with how the 5 man lineup that is on the court performs relative to their opponent's lineup; that is how you win more games.

As Phil Jackson said, the strength of each player is the team. This study removes the noisy obsession the NBA and media have for individual statistics and instead focuses on the success of the entire lineup as the foundation for player rankings. To provide a blunt example, if lineups tend to have higher offensive ratings when player A is on the court compared to when a different player B is on the court, it is logical to say that player A has a larger positive impact on how well efficiently his lineups score.

In summary, this study will attempt determine which players are most influential to the success of the lineups they were apart of through the lens of network analysis. Network analysis is a subset of graph theory analyzing a graph where the edges have a significant situational meaning. There will be two stages in the network creation phase of this study.

The first will create bipartite networks – networks where nodes are divided into two disjoint sets – where there will be one set of "lineup" nodes and one set of "player" nodes for each team in the NBA. The second stage will be projecting these bipartite networks into a unimodal, player-only network.

The influence rankings will be derived from these weighted, player node-only networks by determining eigenvector centrality of each player node. In simple terms, eigenvector centrality measures a node's importance while giving consideration to the importance of its neighbors. Mathematically, this centrality score is determined by performing a matrix calculation to determine the principal eigenvector using the unimodal network's weighted adjacency matrix – a matrix where rows and columns are assigned to nodes in network and weight of the edge between nodes is symbolized numerically. Fundamentally, the concept of eigenvector centrality score representing a node's influence within the network is essential to the conclusions that spur from this study.

For this study, a lineup's "success" will be determined by how the lineup faired for a specific metric (e.g. Offensive Rating, Net Rating, Field Goal Percentage, etc.) throughout the minutes they played together during the 2018-19 NBA season. Only lineups that played a minimum of 27 minutes together throughout the season are considered to protect from effects due to outliers.

An "influence" ranking is created for the players of each team for multiple different metrics that serve as determinants of success. With these rankings, it can be seen which players increase the success of the lineups they were apart of. To be clear, a player's ranking should not be thought of as representative of how well the player performed individually, but how influential he was to the success of the 5-man unit that was on the court.

A second use case can be derived after the weighted, player-only networks are created. By selecting a single player node to focus from, a ranking of his teammates – the other player nodes – he performs best with by ranking the weights of the edges coming from his player node.

This influence ranking will encompass the immeasurable impacts that a player may have on his teammate's performances. To tie back to the Zen Master's quote, the team's strength is its individual members; if teammates perform at higher level when a player is on the court with them, then that player should be valued greatly. There are certain players who basketball fans call "glue guys" or players who have "a strong locker room presence" or players who "make everyone around them better." Every basketball player, coach or fan has heard these descriptions. These players commonly are thought of as players who posses what NBA scouts and fans call the "intangibles." These might be guys who set good screens or keep good floor spacing or always seem to come up with possession of 50-50 loose balls. The goal of this analysis is to properly value these players, whose contributions might not fit in the box score.

# METHODS

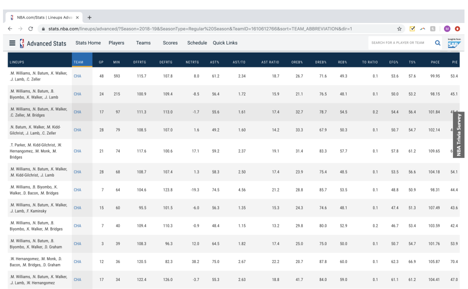
## Acquiring Data

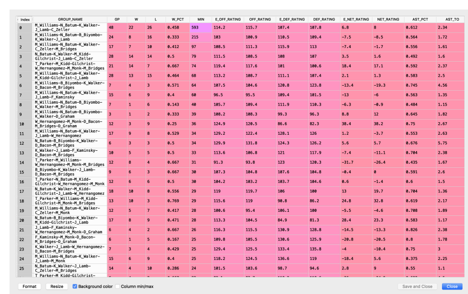
Before any analysis could be done or any networks created, data must be acquired. Reliable, consistent and filterable data is the bedrock for an analysis like this. During each

NBA game, an immense amount of data is collected. The league organizes this data in various ways, including recording data for each 5-man lineup that every NBA team plays. Fortunately, the forward-thinking NBA promotes individual statistical analyses and makes all of the league's game data publicly available through URL endpoints to its stats.nba.com webpage.

This study used the nba_api (https://github.com/swar/nba_api) python package to draw lineup-specific data for each NBA team for the 2018-19 season from the stats.nba.com endpoint (https://stats.nba.com/stats/teamdashlineups).

Below, see a comparison of how the data for the Charlotte Hornets lineups looks on the NBA's stats webpage and as a pandas dataframe:





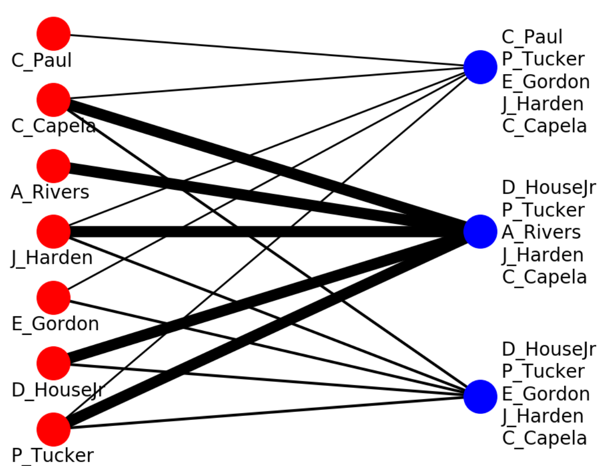
## Building Networks and NetworkX

With the dataframes of lineup specific data available for each team, the NetworkX package was used at both stages of the network creation. NetworkX is a Python package for the study of the structure, dynamics, and functions of complex networks; it contains various modules that can be used to build and manipulate these networks. This package included all modules and functions that were used for this analysis.

### Bipartite Networks

The "bipartite" module is needed to create the first stage networks for each team. At the start of each ranking, the metric that will determine the "success" of the lineup must be selected. The variety of metrics that could be chosen can be seen by looking at the lineup statistics in each team's dataframe. It is important to consider the desired conclusions the rankings should lead to when choosing this metric. For example, if it is desired to see which players make lineups shoot better from the 3 point range, then the three-point field goal percentage should be chosen as the metric that defines success. This metric value as well as the number of minutes the 5-man unit played together are obtained for each lineup node. To increase efficiency during the projection step (described later), the relevant information are stored as attributes of the lineup node.

An example subgraph of the Houston Rockets team bipartite graph with offensive rating as the defining metric is displayed below. The width of each edge is representative of that lineup's metric value:



### Projection to Player Only Networks

The projection step to the unimodal, player-only networks is performed with generic_weighted_projected_graph function in the bipartite module of the NetworkX package.

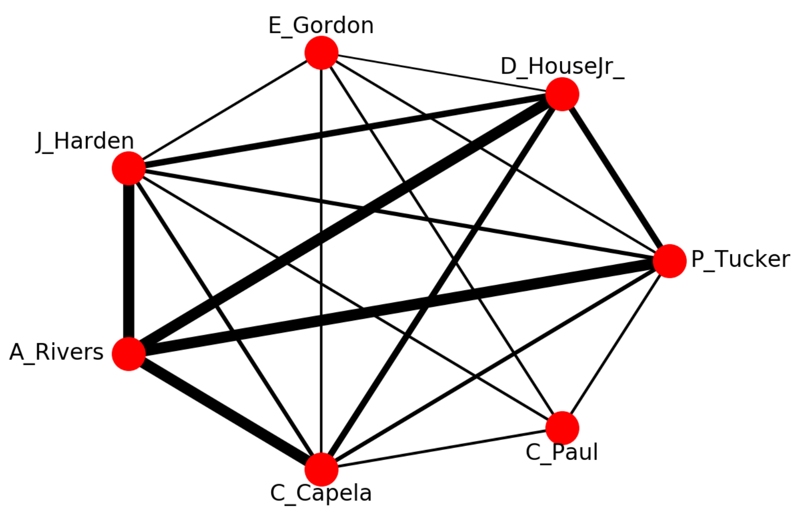
This step creates another network for each team where only players who participated in qualifying lineups are represented. Each player that shared at least one lineup is connected by an edge in this player-only network. The weight of each edge is the weighted average of the two player's shared lineups success with respect to how many minutes each lineups played. This weighted average presents a more representative picture of how well the two players actually performed together compared to adding the metrics values of shared lineups or an unweighted average of the metric values.

$$EdgeWeight_i = \frac{\sum(MinutesPlayed_i * MetricValue_i)}{\sum MinutesPlayed_i} \tag{1}$$

To provide a simple example as to why the projected network's edge weights should be an average rather than a sum, say player A and player B shared 3 lineups and player A and player C shared 6 lineups. For simplicity, say all 9 of these lineups recorded the same metric value. Therefore, player A plays equivalently well when player B or player C is on the court. However, if the edge weight between players is merely a sum of the metric values, then the edge between player A and player C will be weighted twice as heavily as the edge connecting player A and player B.

To provide a simple example as to why the average needs to be weighted by minutes played, say player A and player B shared 2 lineups in the 2018-19 season. One of the lineups recorded 100 minutes and the second lineup only played 10 minutes. Say the first lineup recorded a metric value of 95 over the 100 minutes and the second lineup only recorded a metric value of 75 over the 10 minutes. If the edge weight between player A and player B in player-only graph was determined by an unweighted average of the shared lineup metric values, the weight would be 85 ($(95 + 75)/2)$). Intuitively the first lineup is more representative of the interaction of player A and player B since its metric value was earned over a greater sample size. With a weighted average, the edge weight would be 93.18 ($(10 * 75 + 100 * 95)/(10 + 100)$). This is much closer to the first lineup's metric value, as desired. Obviously, these are two exaggerated cases but the intuition holds true for the real data.

Below, see the projection of the example subgraph of the Houston Rockets bipartite network. The edge's thickness is representative of the edge's weight:




**Creating the Rankings**

As described in the Introduction, after the weighted, player-only networks are created, eigenvector centrality scores are calculated for each player node. The higher the centrality score, the greater the influence that node has on the team's network. Since the edge weights represent a weighted average of the metric value of shared lineups, a greater centrality score reflects influencing the metric to increase. So if it is a metric like offensive rating, a player with the highest centrality score influences the lineups he is apart of to score more points per 100 possessions (which is what offensive rating measures) than any other player on the roster.

The teammate rankings do not involve any node centrality calculations. Therefore, when looking at player A's teammate rankings, it is a not a ranking of which teammates are most influential in improving the performance of player A. The ranking is simply ranking which player nodes are connected to player A with the heaviest weighted edge. In other words, it is a ranking of how successful the lineups each player shared with player A were. Thus, the player who is connected to player A with the most heavily weighted edge is the player that player A plays the best with.

# RESULTS

With so many possibilities for which metric to define lineup success, there are a vast amount of rankings that could be created. A few interesting examples with brief explanations as to why they were selected are included below.

**Golden State Warriors - True Shooting %**

Ranked in Descending Order of Influence

1. Klay Thompson
2. Stephen Curry
3. Andre Iguodala
4. Kevin Durant
5. Draymond Green

6. Jonas Jerebko

7. Kevon Looney

8. Sean Livingston

**Explanation:** This was an noteworthy example because the success of Golden State Warriors dynasty of the mid 2010s was grounded in their profound shooting abilities. True Shooting Percentage is a weighted field goal percentage that accounts for the value of the shot being attempted and was the metric that defined how successful a lineup was in this example. This ranking example shows Andre Iguodala ranked above Kevin Durant, who is unarguably a better pure shooter than Iguodala. This is interesting as it shows that Andre Iguodala is able to improve the True Shooting of his lineups without necessarily being the best shooter; in other words, Iguodala's presence on the court helps the 5-man unit shoot more efficiently more than Durant's. To basketball fans, this type analysis favoring Iguodala may not be too surprising as Iguodala embodies the role of a "glue guy" for the Warriors that was described in the Introduction. Kevin Durant, while an elite isolation scoring and generational offensive talent, plays a role in the offensive that is not as implicit in the success of his teammates.

**Toronto Raptors - Offensive Rating**

Ranked in Descending Order of Influence

1. Fred VanVleet

2. Pascal Siakam

3. Danny Green

4. Kyle Lowry

5. OG Anunoby

6. Kawhi Leonard

7. Serge Ibaka

8. Jonas Valanciunas

**Explanation:** This example is included because it shows how some rankings do not match common expectation. The Toronto Raptors most talented player in 2018-19 was Kawhi Leonard, yet he ranks 6th on this offensive rating influence ranking. For most rosters, the player who is commonly thought of as the top offensive talent was ranked first when offensive rating was used to define success (Damien Lillard for the Trail Blazers, Bradley Beal for the Wizards, Kyrie Irving for the Celtics, etc.) One unique element of Kawhi's season in Toronto was his rest schedule due to prior injuries. Kawhi missed 22 games and played limited minutes in many others. More importantly, due to his limited availability, Kawhi's minutes were likely played disproportionately against higher level of talent. To clarify, if a top-tier player like Leonard can only play 20 minutes in a game the coach will make sure to allocate those 20 minutes when Raptors needed him most (i.e. when the other team was playing their best players). Additionally, it is intuitive that a player on a load management scale would be more likely to sit out for rest against less talented teams. This logic – which claims that Kawhi's lineups play more minutes against higher levels of competition – is a possible explanation for the all-star's low ranking.

These rankings are also interesting because 24 games into the 2019-20 NBA season, Fred VanVleet and Pascal Siakam are having "breakout" years. The Raptors have been a surprise team that the media largely wrote off after the Leonard's departure. These rankings show that because of VanVelet and Siakam's influence last year the basketball community should not be as shocked by Toronto's start this year.

**Milwaukee Bucks - Offensive Rating vs. Net Rating**

Ranked in Descending Order of Influence

| Offensive Rating | Net Rating |
|---|---|
| 1. Khris Middleton | 1. Eric Bledsoe |
| 2. Malcolm Brogdon | 2. Giannis Antetokounmpo |
| 3. Eric Bledsoe | 3. Brook Lopez |
| 4. Giannis Antetokounmpo | 4. Nikola Mirotic |
| 5. Ersan Ilyasova | 5. Tony Snell |
| 6. Brook Lopez | 6. Khris Middleton |
| 7. Tony Snell | 7. Pat Connaughton |
| 8. George Hill | 8. Malcolm Brogdon |

**Explanation:** This example shows the rankings for the Milwaukee Bucks with two different metrics that define success. First, offensive rating, which as described earlier, measures how many points are scored per 100 possessions. Second, net rating, which measures the lineup's point differential per 100 possessions ( = offensive rating - defensive rating).

With one metric just valuing the lineup's offensive success and one measuring success on both ends of the floor, it makes sense the rankings are not identical. Giannis Antetokounmpo won the Most Valuable Player Award for the 2018-19 season, but was also in the discussion for Defensive Player of the Year award. With defensive rating included in the definition of success, it makes sense that Giannis raises in influence as his length, speed and ability to switch and guard multiple positions make his lineups to excel on the defensive end of the floor. Similarly, Brook Lopez jumps in influence in this comparison when defense is factored in. Lopez had a standout defensive year and established himself an elite rim protector. With

Lopez on the court, other Bucks could play more aggressively on the defensive end knowing the 7-footer would be able to provide help defense if they got beat.

**Devin Booker Teammates - Offensive Rating**

Ranked in Descending Order of Shared Edge Weight

1. Kelly Oubre Jr
2. Elie Okobo
3. Tyler Johnson
4. Mikal Bridges
5. Jamal Crawford
6. DeAndre Ayton
7. De'Anthony Melton

**Explanation:** This is an example of the second use case with all-star combo guard Devin Booker being the focal point. It is interesting to see that only one big man (DeAndre Ayton) ranked in the top 7 teammates with offensive rating defining success. Devin Booker is one of the best pure scorers in the NBA. Booker has excelled as both an offensive's main ball handler and as an off-ball threat. These rankings show that he shares more offensively successful lineups with other versatile wing/guards than with big men, who typically restrict floor spacing. As someone who can score at the rim, from midrange and from beyond the three-point line, it makes sense that Booker's best lineups include teammates who space the floor and can also handle the ball.

# CONCLUSION

As described in the introduction, the goal of this study was to investigate how valuable each player is to their team. The goal of every team in the NBA is to win more games and win more championships. In order to achieve these, front offices must be able to create a roster that optimizes talent within the confines of the salary cap and coaching staffs must properly allocate minutes to players on his roster. With a more holistic understanding of how influential players on their roster are to the success of the team, coaches can better tailor their minutes allocation to favor a specific game plan. Additionally, by analyzing which players the team's centerpiece player performs best with, general managers can better target players around the league to pair with their current talent.

## Player Rankings

Each head coach thoroughly examines his opponent's strengths and weaknesses before each game. Naturally, the coach will want to capitalize on his opponents weaknesses and minimize the impact of their strengths. After identifying these strengths and weaknesses, understanding which of his players are most influential in these areas could play an important role in his gameplan.

For example, the Dallas Mavericks and Atlanta Hawks placed in the bottom two spots for transition defense, allowing the most fastbreak points in the NBA. To capitalize on this weakness, a head coach may want to try to increase the pace of the game (i.e. more fastbreak possessions and less half court offensive sets). Erik Spoelstra, the head coach of the Miami Heat, could look at the influence rankings with lineup pace as the metric that defines pace to figure out which players should get more minutes to suit this game plan. Interestingly, Kelly Olynick ranks higher than other Heat big men Bam Adebayo and Hassan Whiteside. While Olynick may be less athletic and slower on an individual comparison basis, the rankings show that he is more influential in increasing the pace of the game; thus, Spoelstra could increase Olynick's minutes against the Mavericks and the Hawks.

Similarly, the Oklahoma City Thunder caused the most turnovers in the league during the 2018-19 season. When the Heat match up against the Thunder, Spoelstra could look at the rankings built with the lineup's assist-to-turnover ratio defining success. Looking at the rankings, Spoelstra would see James Johnson and Tyler Johnson – two players who averaged over 20 minutes a game for the Heat – ranked very low, and might decide to play Josh Richardson and Justice Winslow more minutes than usual as they both rank in the top 4 for this metric.

**Teammate Rankings**

While head coaches could also use the teammate rankings to tailor minutes allocation around a single player, an alternative use case for these is in the front office. The NBA is often called a "top-down" league. This means that teams are built around top-tier talent because a single player can impact the game so severely. Obviously, in today's NBA, no one player can carry his team to a championship, and in order for a team to be a contender they need at least one "superstar" level player, if not more. A defining feature of a team's front office is how efficiently they build around their top tier talents.

Whether a team acquires a superstar through free agency or a trade or a recent draft pick shows enough promise to develop into a potential superstar, front offices must make personnel moves to optimize their best players. Assuming a general manager has already decided a player on his roster can be the centerpiece to a championship team, developing a better picture of which players he plays his best with is key to properly building around him.

The Utah Jazz are in the position of building around a budding superstar with their third-year guard Donovan Mitchell. Since his rookie year, the combo of a young Mitchell and all-star center Rudy Gobert looked to be a solid, young foundation. However, when looking at Mitchell's teammate rankings (with offensive rating as the defining metric), the importance of sharp-shooting forward Joe Ingles is exposed. The lineups that Mitchell and Ingles share are more successful offensively than any other of Mitchell's partnership. This is why signing Ingles to a contract extension before the start of the 2019-20 season was a great move. If they were to let Ingles go in the future, they should prioritize replacing him with another wing who can shoot a high percentage from three.

The Sacramento Kings also have a rising star at point guard that was also selected in the 2017 draft, De'Aaron Fox. Unlike Mitchell, Fox has no other all stars on his roster. The Sacramento front office can start the process of building around Fox from scratch. By understanding which types of players on the Kings roster he currently thrives with, the general manager can better target player's to acquire or draft to pair with him. Fox's teammate ranking (with offensive rating as the defining metric) shows that Fox performs well with big men; three of the top four rankings are held by players over 6'10", but none of which are top tier talents. Based on this analysis, the first step to build around Fox would be to find a high-caliber front court player.

## Further Work

The proper way to value the holistic contributions a player makes to his team will never be found exactly. While this study offers a method to solve this problem, there are obviously improvements and further work that can be done.

First, there are many more metrics available through the nba_api to define the success of lineups. For each metric, a ranking for each team can be created. Exploring these various metrics and understanding how player's influence changes is an obvious extension of this study.

Another continuation of this study is to look at data from different years. Looking at one player's influence on his team increases or decreases over a time period would be an interesting extension. Obviously, roster changes over the time period will affect the other names listed but a general idea of how the player has progressed over the time period can be obtained. It would also be interesting to look at how a specific player's teammate rankings change from year to year as they develop different parts of their game.

# REFERENCES

1. Piette J, Anand S, Pham L. Evaluating Basketball Player Performance via Statistical Network Modeling. In: MIT Sloan Sports Analytics Conference; 2011. Available from: http://www.sloansportsconference.com/?p=2840.

2. Sohum Misra. Player Centrality on NBA Teams. Stanford University SCPD. Available from:
http://snap.stanford.edu/class/cs224w-2017/projects/cs224w-88-final.pdf

3. “NetworkX.” NetworkX, Oct. 2019, networkx.github.io/.

4. “NBA Stats.” NBA Stats on TeamRankings.com, www.teamrankings.com/nba/stat/f.

5. Golbeck, Jennifer. “Eigenvector Centrality.” Eigenvector Centrality - an Overview — ScienceDirect Topics, www.sciencedirect.com/topics/computer-science/eigenvector-centrality.

6. “Graph Theory: Adjacency Matrices.” Https://Www.ebi.ac.uk, European Bioinformatics Institute (EMBL-EBI), www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/introduction-graph-theory/graph-0.

7. “Basketball Statistics and History.” Player Stats, www.basketball-reference.com/.